PEPTIDESTRUCTURE

Table of Contents

FUNCTION

[Top | Next]

PeptideStructure makes secondary structure predictions for a peptide sequence. The predictions include (in addition to alpha, beta, coil, and turn) measures for antigenicity, flexibility, hydrophobicity, and surface probability. PlotStructure displays the predictions graphically.

DESCRIPTION

[Previous | Top | Next]

PeptideStructure makes predictions of the following features of an amino acid sequence:

```
-  Secondary structure according to the Chou-Fasman method
-  Secondary structure according to the Garnier-Osguthorpe-Robson method
-  Hydrophilicity according to either the Kyte-Doolittle or Hopp-Woods
 method
-  Surface probability according to the Emini method
-  Flexibility according to the Karplus-Schulz method
-  Glycosylation sites
-  Antigenic index according to the Jameson-Wolf method
```

The results are written into a file for graphical presentation with PlotStructure.

PeptideStructure uses the original method of Chou and Fasman (Adv. in Enzymol., **47**; 45-148 (1978)) to predict helices, sheets, and turns. It resolves overlapping regions of alpha-helices and beta-sheets with the *overall probability* procedure introduced by K. Nishikawa (Biochim. Biophys. Acta, **748**; 285-299 (1983)). This same procedure also locates turns that are not in conflict with other secondary structures. The Chou-Fasman rules are slightly modified as follows: *Helix:* the condition that p(bound) be greater than 1.0 and that p(alpha) be greater than p(beta) are not used; and *Sheet:* a minimum length of five residues is required.

PeptideStructure also predicts secondary structure according to a slightly modified method of Robson-Garnier (Garnier et al. J. Mol. Biol., **120**; 97 (1978)). The minimum length of an alpha-helix is six and of a beta-sheet is four. Regions without adequate predictions are replaced by the conformational state of the next best probability.

Hydrophilicity is calculated according to the algorithm of Kyte and Doolittle (J. Mol. Biol., **157**; 105-132 (1982)). The window, normally set to seven residues, can be changed with **-HWIN**dow.

Surface probability is calculated according to a formula of Emini et al. (J. Virol., **55(3)**; 836-839 (1985)), which is slightly modified for the end values of the protein chains. The single probabilities are taken from Janin et al. (J. Mol. Biol., **125**; 357-386 (1978)).

Glycosylation sites are predicted for sites where the residues have the composition NXT or NXS. When X is D, W, or P, the site is taken to be a weak glycosylation site, otherwise it is a strong glycosylation site.

The antigenic index (AI) is a measure of the probability that a region is antigenic. It is calculated by summing several weighted measures of secondary structure. The method is described by Jameson and Wolf (CABIOS, **4(1)**; 181-186 (1988)) and the formula is defined under the ALGORITHM topic below. Peaks of antigenic index can be smoothed with **-BROA**dening.

EXAMPLE

[Previous | Top | Next]

Here is a session using PeptideStructure to calculate the secondary structure of the outer membrane protein F precursor from E. coli (PIR:Mmecf).

```
% peptidestructure

 PEPTIDESTRUCTURE for what peptide sequence ?  PIR:Mmecf

                  Begin (* 1 *) ?
                End (*   362 *) ?

 Calculate hydrophilicity according to

     H)opp-Woods or
     K)yte-Doolittle

 Please choose one (* K *) :

 What should I call the output file (* mmecf.p2s *) ?

%
```

OUTPUT

[Previous | Top | Next]

Here is part of the output file:

```
PEPTIDESTRUCTURE of: pir1:Mmecf  check: 2147  from: 1  to: 362

outer membrane protein F precursor - Escherichia coli

Hydrophilicity (Kyte-Doolittle) averaged over a window of: 7
Surface Probability according to Emini
Chain Flexibility according to Karplus-Schulz
Secondary Structure according to Chou-Fasman
Secondary Structure according to Garnier-Osguthorpe-Robson
```

```
Antigenicity Index according to Jameson-Wolf

                    Date: October 21, 1998 10:51

Pos  AA  GlycoS  HyPhil  SurfPr  FlexPr  CF-Pred GORPred AI-Ind ..

1     M       .   1.150   1.437   1.000       .       H   0.900
2     M       .   1.620   1.808   1.000       .       H   0.900
3     K       .   0.600   0.991   1.000       .       H   0.750

////////////////////////////////////////////////////////////

360   Y       .  -1.050   0.358   1.000       B       B  -0.450
361   Q       .  -1.340   0.653   1.000       B       B  -0.450
362   F       .  -0.550   1.125   1.000       B       B  -0.300
```

INPUT FILES

[Previous | Top | Next]

PeptideStructure accepts a single protein sequence as input. If PeptideStructure rejects your protein sequence, turn to Appendix VI to see how to change or set the type of a sequence.

RELATED PROGRAMS

[Previous | Top | Next]

PlotStructure plots the results from PeptideStructure. PepPlot plots parallel curves of standard measures of protein secondary structure.

RESTRICTIONS

[Previous | Top | Next]

PeptideStructure requires a peptide sequence. The ambiguous amino acid symbols B (Asx), Z (Glx), and X (unknown), along with the symbol * (termination codon) are not supported for this program (see Appendix III).

ALGORITHM

[Previous | Top | Next]

See the papers cited under the DESCRIPTION topic above. Do *not* attempt to interpret protein secondary structure predictions without reading the Robson-Garnier paper.

The antigenic index (AI) is calculated by summing several weighted measures of secondary structure. The method is described by Jameson and Wolf (CABIOS, **4(1)**; 181-186 (1988)). The formula for the antigenic index (AI) is:

```
AI = 0.3*[H] + 0.15*[S]+ 0.15*[F] + 0.2*[Cs] + 0.2*[Rs]

where:

[H] =                2       1       -1      -2          for
Hydrophilicity       >0.5    >0.     >-0.4   <-0.4

[S] =                1               0                   for
Surface probability  >=1.0           <1.0

[F] =                1               0                   for
```

```
Flexibility          >=1.0           <1.0

[Cs] =                 2               1             0       for
Chou-F.          strong turn     turn or coil   other

[Rs] =                 2               1             0       for
Robson-G.        strong turn     turn or coil   other
```

Peaks of antigenic index can be smoothed with **-BROA**dening. In each case where the antigenic index has a peak (except inside a strong helix), the index can be decremented (broadened) smoothly to at least 80%, 40%, and 20% of the peak in the three surrounding amino acids.

CONSIDERATIONS

[Previous | Top | Next]

You should realize that measures of protein secondary structure are only weakly correlated with actual structures. The Chou-Fasman method was designed to apply to soluble (globular) proteins.

COMMAND-LINE SUMMARY

[Previous | Top | Next]

All parameters for this program may be added to the command line. Use **-CHE**ck to view the summary below and to specify parameters before the program executes. In the summary below, the capitalized letters in the parameter names are the letters that you *must* type in order to use the parameter. Square brackets ([and]) enclose parameter values that are optional. For more information, see "Using Program Parameters" in Chapter 3, Using Programs in the User's Guide.

```
Minimal Syntax: % peptidestructure [-INfile=]pir:mmecf -Default

Prompted Parameters:

-BEGin=1 -END=362           sets the range of interest
-MENu=k                     chooses hydrophilicity method:
                              K=kyte-Doolittle or H=hopp-Woods
[-OUTfile=]mmecf.p2s        names the output file

Local Data Files: None

Optional Parameters:

-HWINdow=7                  sets the window for the hydrophilicity calculation
-BROAdening                 broadens peaks of antigenic index outside
                              of strong helices
-RSF[=peptidestructure.rsf] saves results as features in the RSF file
```

ACKNOWLEDGEMENT

[Previous | Top | Next]

PeptideStructure and PlotStructure were communicated to us by Dr. Hans Wolf for Drs. Susanne Modrow and Manfred Motz of the Max von Pettenkofer-Institut of the University of Munich and for Bradford Jameson of the California Institute of Technology. The programs were written for them by Dr. B. J. Foertsch and G. Herrmann and were modified for compatibility with Version 5 of the Wisconsin Package by John Devereux. The graphic schematic and overlying symbols for hydropathy and glycosylation are based on the early work (1983) of Ellis Golub of the University of Pennsylvania. These programs have been described and used by Cohen et al. (J. Virol.

49; 102-108 (1984)), Starcich et al. (Cell, **45**; 637-648 (1986)), Motz et al. (Gene, **42**; 303-312 (1986)) and by Modrow and Wolf (Proc. Natl. Acad. Sci. USA **83**; 5703-5707 (1986)). The method for calculating antigenic index is described by Jameson and Wolf (CABIOS, **4(1)**, 181-186 (1988)). Any use of PeptideStructure or PlotStructure for published research should cite the Jameson and Wolf article.

LOCAL DATA FILES

[Previous | Top | Next]

None.

PARAMETER REFERENCE

[Previous | Top | Next]

You can set the parameters listed below from the command line. For more information, see "Using Program Parameters" in Chapter 3, Using Programs in the User's Guide.

-MENu=K

sets the hydropathy scale to use: Kyte-Doolittle (K) or Hopp-Woods (H).

-HWINdow=7

lets you set the window of integration for the hydrophilicity measurement. The window is usually set to seven residues.

-BROAdening

sets PeptideStructure to broaden the peaks of the antigenic index as described under the ALGORITHM topic above.

-RSF=peptidestructure.rsf

writes an RSF (rich sequence format) file containing the input sequences annotated with features generated from the results of PeptideStructure. This RSF file is suitable for input to other Wisconsin Package programs that support RSF files. In particular, you can use SeqLab to view this features annotation graphically. If you don't specify a file name with this parameter, then the program creates one using peptidestructure for the file basename and .rsf for the extension. For more information on RSF files, see "Using Rich Sequence Format (RSF) Files" in Chapter 2 of the User's Guide. Or, see "Rich Sequence Format (RSF) Files" in Appendix C of the SeqLab Guide.

Printed: January 9, 2002 13:45 (1162)

[Program Manual | User's Guide | Data Files | Databases]

Technical Support: support-us@accelrys.com
or support-eu@accelrys.com

www.accelrys.com/bio